

César Augusto Bernal Campos

Palabras Claves - Keywords

Web semántica, Web Sintáctica, RDF, XML, MARC21, Estándares, Meta datos, Ontologías.

1. Introducción

El presente artículo describirá de manera breve la conceptualización de lo que es la Web semántica y sintáctica; propondrá unas preguntas que ocupan al profesional en ciencias de la información, logrando en las respuestas de las mismas hallar o no, la importancia de lo que se ha denominado como la Web Semántica, para luego sacar conclusiones en el contexto de lo que se esta hablando.

El proyecto de la Web Semántica ha sido impulsado por el World Wide Web Consortium (W3C), dentro de su dominio de actividades denominado Technology & Society. Puede considerarse, además, un proyecto personal del inventor de la Web y fundador y director del W3C, Tim Berners Lee.ⁱ Como componente esencial de este proyecto la W3C lo enmarca dentro de la tecnología y la sociedad, apuntando desde su comienzo que para este proyecto la sociedad es importante y para ella es el avance que pueda tener la misma en la aplicación de los estándares, para la accesibilidad de los documentos y de los objetos que se encuentran en la Web.

2. En que se basa la Web semántica

Se basa en RDF acrónimo que significa *Resource Description Framework*, el cual es un marco de descripción de recursos, es un lenguaje de descripción cuya función principal es describir cómo se representan los datos en las páginas Web (HTML) ecuaciones (MathML, LaTeX, CML), representaciones geográficas (GML), animaciones (SMIL) entre otros, también utiliza SPARQL (SPARQL Protocol and RDF Query Language), que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes datos, y OWL es un mecanismo para desarrollar temas o vocabularios específicos en los que se asocian esos recursos. Lo que hace OWL es proporcionar un lenguaje para definir ontologías estructuradas que pueden ser utilizadas a través de diferentes sistemas.

3. Definiciones del proyecto de la WS

Desde el contexto de la inteligencia artificial (IA) se define como el *conjunto de iniciativas para promover la organización de la Web, en el sentido que sean estructuradas y codificadas son el propósito de que los ordenadores sean capaces de efectuar y razonarⁱⁱ*; el éxito depende de cómo estén codificados los datos y los objetos, por que así la IA será capaz de simular con efectividad a partir del contenido.

La definición desde la visión de procesamiento robusto, *consiste en convertir a la World Wide Web (WWW) en una gran base de datos capas de soportar un procesamiento sistemático y consistente de la informaciónⁱⁱⁱ*, en esta dinámica el conjunto de la base de datos sería todo el contenido de la Web, pero es sistemático en la medida que los resultados generados por los motores de búsqueda devuelvan contenido a partir de los metadatos que tendrían que estar vinculados con los documentos y objetos para que sea realmente procesamiento robusto.

4. Web sintáctica

Para conocer el código fuente de las páginas Web solo basta dar clic sobre el menú desplegable de ver/código fuente de la pagina ó menú ver/código fuente, respectivamente navegadores Mozilla Firefox, Internet Explorer, de esta manera se puede conocer en lenguaje en que están codificadas.

Ejemplo:

```
1 <!DOCTYPE html PUBLIC "-//W3C//DTD HTML
2 <html>
3   <head>
4     <title>Example</title>
5     <link href="screen.css" rel="sty
6   </head>
7   <body>
8     <h1>
9       <a href="/">Header</a>
10    </h1>
11    <ul id="nav">
12      <li>
13        <a href="one/">One</a>
14      </li>
15      <li>
16        <a href="two/">Two</a>
17      </li>
```

Fig. 1. Codificación html

Notemos en la figura 1, las etiquetas h1, h2, h3, etc. donde h1 esta destinada para el nombre de la pagina, y cada una de ellas van representando un nivel dentro de la pagina, hay otras etiquetas que definen las propiedades comunes a toda la pagina como color de fondo, márgenes y estilos, estas son unas a manera de ejemplo: * texto en negrita, <i> texto en cursiva <u> texto subrayado.^{iv}*

5. Procesamiento de información en páginas Web

Los ordenadores construyen índices con las palabras que aparecen en las páginas web, luego lo que hace el ordenador es comparar las palabras de la pregunta que se envía a través del motor de búsqueda en las palabras de su índice.

Ejemplo:



Fig. 2. Búsqueda web semántica en google

En este ejemplo la pregunta es “web semántica” utilizando como motor de búsqueda google, lo que hará el motor de búsqueda es comparar las palabras de la pregunta con las palabras de su índice; al abrir el primero recurso y observar el código fuente de la página como se indicó anteriormente nos encontramos con la siguiente etiqueta:

```
<title>Web semántica - Wikipedia, la
enciclopedia libre</title>
```

Observemos que el valor de la etiqueta es equivalente a la pregunta.

6. La Web semántica

Se espera con esta iniciativa que los sistemas de información sean capaces de buscar por conceptos, esto denota de manera directa lo que intenta la IA como mencionamos con anterioridad “*la IA capaz de simular con efectividad a partir del contenido*” esto quiere decir en un contexto semántico que los ordenadores “entiendan” o que traduzcan las preguntas conceptualmente, en el sentido que cuando se pregunte sobre guerra también los sistemas respondan sobre la paz, se espera así que las operaciones que efectúen estos ordenadores pueden hacerla de manera paralela (granulidad), siendo así, las búsquedas tendrán su propia iniciativa involucrando lenguajes documentales, metadatos y ontologías.

7. Infraestructura de la Web semántica

Tim Berners-Lee diagramó la infraestructura de la Web Semántica en 7 capas o niveles como se muestra en la figura 3:

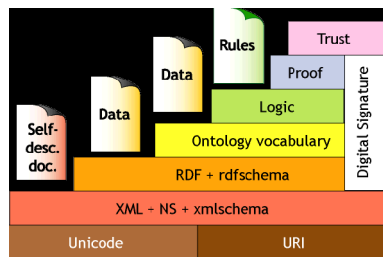


Fig. 3 Infraestructura Web Semantica Tim Berners-Lee

1 nivel de codificación de las páginas.

2 nivel estándar con carga semántica.

3 nivel, categorías de paginas interoperables RDF

4 nivel sistema de ontologías OWL

Se puede apreciar que pasar por estas capas, pretende establecer confianza (Trust), esto se gana en la medida en que se incrementa el nivel, para así llegar a un nivel optimo de la información recuperada.

8. ¿Un Lenguaje XML es valido para las unidades de información?

Si bien XML permite definir tipos de documentos y los conjuntos de las etiquetas necesarias para codificar tales documentos; cómo llevar este lenguaje para la descripción de objetos documentales en el ámbito bibliotecológico y archivístico sería una de las preguntas que el profesional en Ciencias de la información, debería plantearse; encontrando la respuesta, el siguiente planteamiento sería analizar si este lenguaje genera más confianza (Trust) en el momento que el usuario busque y recupere información, esta confianza traducida en términos bibliotecológicos no es otra diferente a la relevancia, entonces dentro de esta óptica tendríamos más preguntas ¿Cuál de los lenguajes documentales es mejor para las unidades de información? ¿Verdaderamente un lenguaje de marcado como XML y el pasar por cada una de las capas de la Web semántica genera confianza (Trust)? Seguramente estas preguntas tienen respuesta, pero lo que queda por indagar y conocer es si ya existen practicas aceptadas y validadas en este contexto en el ámbito bibliotecológico, para decidir si el panorama que se vislumbra dará viraje al intento de normalización que las unidades de información siempre han buscado.

9. RDF y etiquetas Marc21 bibliográfico ¿son comunes?

RDF permite describir recursos; Marc21 bibliográfico es un formato estándar para la representación de registros bibliográficos legibles por máquinas, en donde los registros

bibliográficos de Marc21 son los recursos y en RDF además son los objetos, hasta las personas ya que el número de identificación de las mismas como el número de cedula, el código del carne universitario tiene la misma connotación, ya que estas describen el individuo o lo que en la Web semántica se denomina ontología es decir la especificación de una conceptualización, RDF además de expresar el significado de los recursos no se limita únicamente el documento primario, es decir el original en otras palabras el texto, la revista, el objeto documental en sí, sino que además tiene la misma connotación como documento primario todo los registros que se recuperen bajo la infraestructura de la Web Semántica, dado que este recurso proporciona información y satisface una necesidad de información bajo una pregunta dada; sería apresurado decir que RDF y Marc21 son comunes, pero lo que si queda claro, es que los dos son elementos de descripción para los recursos de información.

10. Un ejemplo de Ontologías

El número de la cedula es un metadato que en RDF es un atributo o una descripción, en este mismo escenario en OWL la clase sería identificación ID, es decir la especificación de una conceptualización, resulta complejo entenderlo de esta manera, para ello se presenta la siguiente figura a manera de ilustración.



Fig. 4 Identificación de RDF, Metadato y OWL

11. Conclusiones

La estructura de la Web que se conoce esta creada sintácticamente.

El proyecto de la W3C persigue estructurar la Web semánticamente.

Los lenguajes de descripción como RDF son un marco de descripción de recursos, cuya función principal es describir la representación de los datos en las páginas Web.

Los recursos que se encuentra en la red publica de la Internet son documentos, objetos, etc., que proporcionan información, pero el éxito de su recuperación radica en como estén estructurados para que den el carácter de relevancia y pertinencia para el usuario, independientemente los motores de búsqueda utilizados.

El profesional de ciencias de la información debe conocer como se estructuran los objetos en la Web para que con ello y en el momento que se incorporen tecnologías en las unidades de información estas sean consecuentes con la internet del futuro.

La descripción, análisis, clasificación de diferentes objetos documentales han de ser etiquetados, para que su búsqueda y recuperación sea de más valor para los usuarios.

La adecuada incorporación y atención por parte del profesional en ciencias de la información en el contexto de la Web semántica como proyecto de la W3C, logrará que si participa en la misma, descubra un nuevo campo de acción de acuerdo a las nuevas tecnologías, lo cual se convertirá en ventaja competitiva a nivel profesional y por ende como componente de valor.

Referencias

RDF [término de búsqueda: Define RDF]. [en línea – html]. [consulta: 2008-02-10] Disponible en http://es.wikipedia.org/wiki/Resource_Description_Framework

HTML [término de búsqueda: Define HTML]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://es.wikipedia.org/wiki/HTML>

MathML, LaTeX, CML [término de búsqueda: Define]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://es.wikipedia.org/wiki/MathML>

(SMIL) [término de búsqueda: Define]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://es.wikipedia.org/wiki/SMIL>

SPARQL. [término de búsqueda: Define]. [en línea –html]. [consulta: 2008-02-10] Disponible en <http://www.w3c.es/2001/sw/DataAccess/>

OWL. término de búsqueda: Define]. [en línea –html]. [consulta: 2008-02-10] Disponible en <http://w3.org/2004/OWL/>

Figura 1. [término de búsqueda: h1 h2 HTML]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://es.wikipedia.org/wiki/HTML>

Figura 2. [término de búsqueda: Web semántica]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://www.google.com.co/search?hl=es&q=web+semantica&btnG=Buscar&meta=>

Granulidad. [término de búsqueda: Define granulidad]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://www.powow.com/ireti2000/Granulidad.htm>

Metadatos. [término de búsqueda: Define metadatos]. [en línea – html]. [consulta: 2008-02-10] Disponible en http://www.google.com.co/url?sa=X&start=1&oi=define&q=http://sig.a.cna.gob.mx/SIGA/Diccionarios/glosario.htm&usg=AFQjCNFkQyJ07ixSP4cIgliHwD1fitP_FA

Ontologías. [término de búsqueda: Define Ontologías]. [en línea – html]. [consulta: 2008-02-10] Disponible en [http://www.google.com.co/url?sa=X&start=6&oi=define&q=http://es.wikipedia.org/wiki/Ontolog%C3%ADa+\(inform%C3%A1tica\)&usg=AFQjCNHwgBjST3YrEJP_zpHxJWMmezbZhg](http://www.google.com.co/url?sa=X&start=6&oi=define&q=http://es.wikipedia.org/wiki/Ontolog%C3%ADa+(inform%C3%A1tica)&usg=AFQjCNHwgBjST3YrEJP_zpHxJWMmezbZhg)

Tim Berners-Lee. [término de búsqueda: Tim Berners Lee]. [en línea – html]. [consulta: 2008-02-12] Disponible en <http://www.w3.org/People/Berners-Lee/>

Fig. 3 Infraestructura Web Semantica Tim Berners-Lee. [término de búsqueda: Infraestructura Web semántica]. [en línea – html]. [consulta: 2008-02-12] Disponible en <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

Marc21. [término de búsqueda: Marc21 LOC]. [en línea – html]. [consulta: 2008-02-12] Disponible en http://www.google.com.co/url?sa=t&ct=res&cd=1&url=http%3A%2F%2Fwww.loc.gov%2Fmarc%2Fmarcspace.html&ei=jGKyR7WaGp6geNW1rf8C&usg=AFQjCNE4N97eF7WsApKZPNsLsVwksv0qFw&sig2=U4fNcoyao3W0oF_5C0F-dA



Los contenidos de este documento esta bajo una licencia Creative Commons si no se indica lo contrario

ⁱ Codina, Lluís y Rovira, Cristófol. Recursos sobre la Web Semántica. [término de búsqueda: Web Semántica Rovira Codina]. [en línea - .doc]. [consulta: 2008-02-10] Disponible en http://www.google.com.co/url?sa=t&ct=res&cd=2&url=http%3A%2F%2Fwww.lluiscodina.com%2Fwebsemanticav2.doc&ei=xYe4R_KDEIXOelvyjeMM&usg=AFQjCNGqscO9JRIXMsfqTr6gpk_OII0DLw&sig2=F1V9FssU5F5jNFIUma-tAw

ⁱⁱ Codina, Lluís y Rovira, Cristófol. La Web Semántica. [término de búsqueda: Web Semántica Rovira Codina]. [en línea - .pdf]. [consulta: 2008-02-10] Disponible en http://eprints.rclis.org/archive/00008637/01/web_semantica_.pdf

ⁱⁱⁱ Ibid., p.2

^{iv} Nivel HTML. [término de búsqueda: Niveles HTML]. [en línea – html]. [consulta: 2008-02-10] Disponible en <http://es.wikipedia.org/wiki/HTML>

^v Codina, Lluís R Rovira, Cristofol. Op.Cit., p. 7